Dis6 Info Theory

Wednesday, February 24, 2021 12:11 PM

Agenda

1. Motivations: what question(s) can info theory help us answer?
2. Relationship diagram between key quantities
3. Problems

① Find min # bits on av. to send msg to someone

$H(X)$

4 letter language : A, B, C, D

'00' '01' '10' '11' av. 2bits

What if we know, AACAABDAAA

A = 0.8
B = 0.1
C = 0.05
D = 0.05

A = '0'
B = '1'
C = '10'
D = '100'

0100
ABAA or AD

Source Coding Theorem

a bunch of symbols, eg A, B, C, D, E, ...

mapping → bit sequences

$X_1, X_2, \dots, X_n \sim \text{IID } P_X$, $l(x)$:= Length of x (in bits)

∃ a source coding scheme, s.t.

$\forall \epsilon > 0 \quad \lim_{n\to\infty} E\left[\frac{1}{n} l(X_1, \dots, X_n)\right] \leq H(X) + \epsilon$

$H(X)$

$\forall \epsilon > 0, \quad P(|X_n - X| > \epsilon) \leq \frac{Var(X)}{\epsilon}$

$H(X) = E\left[\log \frac{1}{P_X(X)}\right] = \sum_x P_X(x) \log \frac{1}{P_X(x)}$

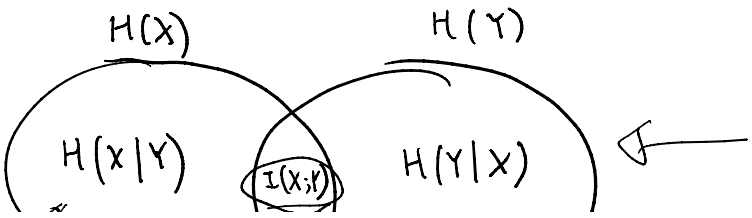
$P_X(x) = 0.002$

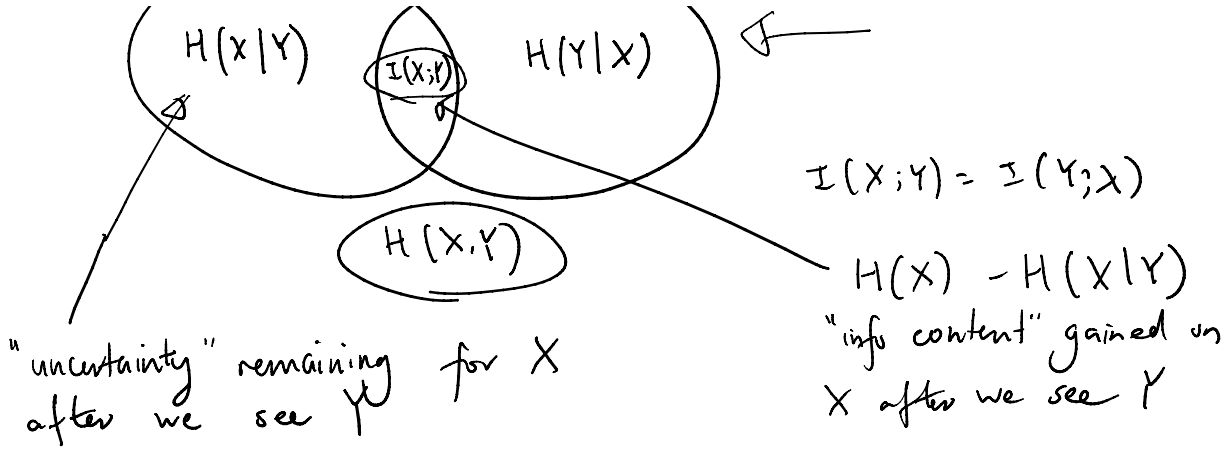
$\frac{1}{P_X(x)}$ = really big

$\log \frac{1}{P_X(x)}$

Expected 'surprise'
'uncertainty'
'info content'

$H(X), \quad H(X,Y), \quad H(X|Y), \quad I(X;Y)$





1. **Mutual Information and Noisy Typewriter**

The **mutual information** of X and Y is defined as

$$I(X;Y) := H(X) - H(X \mid Y)$$

Here, $H(X \mid Y)$ denotes the **conditional entropy** of X given Y, which is defined as:

$$\begin{aligned} H(X \mid Y) &= \sum_{y \in \mathcal{Y}} p_Y(y) H(X \mid Y = y) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y) \log_2 \frac{1}{p_{X|Y}(x \mid y)} \\ &= - \sum_{y \in \mathcal{Y}, x \in \mathcal{X}} p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_Y(y)} \end{aligned}$$

The interpretation of conditional entropy is the average amount of uncertainty remaining in the random variable X after observing Y. The interpretation of mutual information is therefore the amount of information about X gained by observing Y.

(a) Show that $H(X,Y) = H(Y) + H(X \mid Y) = H(X) + H(Y \mid X)$. This is often called the **Chain Rule**. Interpret this rule.

$$H(X|Y) = E\left[\log \frac{1}{P_{X|Y}(X|Y)}\right]$$

$$H(X,Y) = E\left[\log \frac{1}{P_{X,Y}(X,Y)}\right]$$

$$= E\left[\log \frac{1}{P_{X|Y}(X|Y)} + \log \frac{1}{P_Y(Y)}\right]$$

$$= E\left[\log \frac{1}{P_{X|Y}(X|Y)}\right] + E\left[\frac{1}{P_Y(Y)}\right]$$

$$= H[X|Y] + H(Y)$$

(b) Show that $I(X;Y) = H(X) + H(Y) - H(X,Y)$. Note that this shows that $I(X;Y) = I(Y;X)$, i.e., mutual information is symmetric.

$$I(X;Y) = H(X) - H(X|Y) = H(X) - \left(H(X,Y) - H(Y)\right)$$

$$= H(X) + H(Y) - H(X,Y)$$

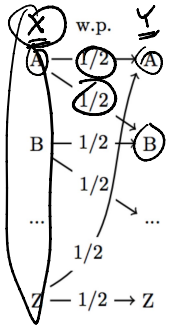
$= H(X) + H(Y) - H(X,Y)$

$= H(Y) - H(Y|X)$

$I(X;Y) = H(X) - H(X|Y)$

~~$H(X,Y) = H(X|Y)$~~

(c) Consider the noisy typewriter.



$P(X = A) = ?$

$P(X = B) = ?$

Each symbol gets sent to one of the adjacent symbols with probability 1/2. Let X be the input to the noisy typewriter, and let Y be the output (X is a random variable that takes values in the English alphabet). What is the distribution of X that maximizes $I(X;Y)$?

Note: distribution that maximizes the entropy of a random variable is the uniform distribution

$I(X;Y) = H(X) - H(X|Y)$

$= H(Y) - H(Y|X)$, $H(Y|X) = 1$

given X, we know

$Y = chr(X)$ w.p. $\frac{1}{2}$

$= chr(X+1)$ w.p. $\frac{1}{2}$

ans. Let X be uniform on $\{A, B, \ldots, Z\}$

$Z \sim Bern(\frac{1}{2})$

$H(Z) = \sum_z P_Z(z) \log \frac{1}{P_Z(z)}$

$= \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$

$= \frac{1}{2} + \frac{1}{2} = 1$

$Y|X \sim Bern(\frac{1}{2})$

$X = A$ w.p. $\frac{1}{26}$ ... Z w.p. $\frac{1}{26}$ $\Rightarrow Y =$ A w.p. $\frac{1}{26}$ ... Z w.p. $\frac{1}{26}$

$\Rightarrow I(X;Y)$ maximised

Note
It turns out that $I(X;Y) \geq 0$ with equality if and only if X and Y are independent. The mutual information is an important quantity for channel coding.

2. Entropy of a Sum

(a) Let X_1, X_2 be i.i.d. Bernoulli(1/2) (fair coin flips). Calculate $H(X_1 + X_2)$ and show that $H(X_1 + X_2) \geq H(X_1)$

$X_1 + X_2 \sim Bin(2, \frac{1}{2})$

$Z = X_1 + X_2 = \begin{cases} 0 & \text{w.p. } \frac{1}{4} \\ 1 & \text{w.p. } \frac{1}{2} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases}$

$X_1 + X_2 \sim \text{Bin}(2, \frac{1}{2})$ $\quad Z = X_1 + X_2 = \begin{array}{|c|c|} \hline 1 & \frac{1}{2} \\ 2 & \frac{1}{4} \\ \hline \end{array}$

$$H(X_1 + X_2) = \sum_z P_X(z) \log \frac{1}{P_X(z)}$$

$$= \frac{1}{4} + \log_2 4 + \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$$

$H(X_1) = 1$ (by formula)

(b) It turns out that in general adding independent random variables increases entropy. To prove this, first let's prove that conditioning decreases entropy. Let X and Y be two (not necessarily i.i.d) random variables, prove that $H(X|Y) \leq H(X)$.
Hint: Use the the fact that $I(X;Y) \geq 0$ for any r.v.s X and Y.

$I(X;Y) = H(X) - H(X|Y) \geq 0$

$\| \quad 0$

(c) Now suppose X_1, X_2 are two independent (but not necessarily identically distributed) random variables. Prove that $H(X_1 + X_2) \geq H(X_1)$.

$$H(X_1 + X_2) \geq H(X_1 + X_2 | X_2) = H(X_1 | X_2) = H(X_1)$$

$$= \sum_{z, x_2} P_{X_1 + X_2, X_2}(z, x_2) \log \frac{1}{P_{X_1 + X_2 | X_2}(z | x_2)}$$